# N741: Homework 5

*Valerie Mac, PhD.*

*March 14, 2017*

# Homework 5 - DUE March 15, 2017

For this homework, we'll work with the "Wong" dataset built in to the `car` package. The "Wong" data frame has 331 row and 7 columns. The observations are longitudinal data on recovery of IQ after comas of varying duration for 200 subjects. The data are from Wong, Monette, and Weiner (2001) and are for 200 patients who sustained traumatic brain injuries resulting in comas of varying duration. After awakening from their comas, patients were periodically administered a standard IQ test, but the average number of measurements per patient is small (331/200 = 1.7). *To get more info type `??Wong` .*

The 7 variables in the dataset are:

- `id`
  - patient ID number.
- `days`
  - number of days post coma at which IQs were measured.
- `duration`
  - duration of the coma in days.
- `sex`
  - a factor with levels Female and Male.
- `age`
  - in years at the time of injury.
- `piq`
  - performance (i.e., mathematical) IQ.
- `viq`
  - verbal IQ. ## Install needed packages

```
install.packages("car",repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/gneis/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)
```

```
## package 'car' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\gneis\AppData\Local\Temp\Rtmpy8Vfhj\downloaded_packages
```

```
install.packages("dyplr",repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/gneis/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)
```

```
## Warning: package 'dyplr' is not available (for R version 3.3.2)
```

```
install.packages("DT",repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/gneis/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)
```

```
## package 'DT' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\gneis\AppData\Local\Temp\Rtmpy8Vfhj\downloaded_packages
```

# Load dataset in from `car` package

```
library(car)
```

```
## Warning: package 'car' was built under R version 3.3.3
```

```
data(Wong)

library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.3.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:car':
##
##     recode
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# add an age group variable
Wong$agegrp <- case_when(
  (Wong$age > 0 & Wong$age <= 10) ~ 1,
  (Wong$age > 10 & Wong$age <= 20) ~ 2,
  (Wong$age > 20 & Wong$age <= 30) ~ 3,
  (Wong$age > 30 & Wong$age <= 40) ~ 4,
  (Wong$age > 40 & Wong$age <= 50) ~ 5,
  (Wong$age > 50 & Wong$age <= 60) ~ 6,
  (Wong$age > 60 & Wong$age <= 70) ~ 7,
  (Wong$age > 70 & Wong$age <= 100) ~ 8)

# convert to factor, add code levels and labels
Wong$agegrp <- factor(Wong$agegrp,
                  levels = c(1,2,3,4,5,6,7,8),
                  labels = c("Ages 1-10",
                             "Ages 11-10",
                             "Ages 21-10",
                             "Ages 31-10",
                             "Ages 41-10",
                             "Ages 51-10",
                             "Ages 61-70",
                             "Ages 71-100"))
```

Using this dataset, and today's demos complete the following tasks:

1. Make a table of non-parametric statistics (median and IQR) for the number of days and duration grouped by `sex`. You'll be using `summarise()` from the `dplyr` package. For a given variable `x` you'll use `median(x, na.rm=TRUE)`, `quantile(x, 0.25, na.rm=TRUE)`, and `quantile(x, 0.75, na.rm=TRUE)`. Give the table a title using the `caption=` option and update the column names with something nice using the `col.names=` option in the `knitr::kable()` command.

# Making a table of non-parametric statistics (median and IQR) for the number of days and duration grouped by `sex`

```
Wong %>%
    group_by(sex) %>%
    summarise(mediandays= median(days, na.rm=TRUE),
              quantile(days, 0.25, na.rm=TRUE),
              quantile(days, 0.75, na.rm=TRUE),
              medianduration= median(duration, na.rm=TRUE),
              quantile(duration, 0.25, na.rm=TRUE),
              quantile(duration, 0.75, na.rm=TRUE))%>%

knitr::kable(col.names=c("Sex","Median Days","1st Quartile Days","3rd Quartile Days","Median Duration","1
st Quartile Duration", "3rd Quartile Duration"),
                caption="Days and Duration: Nonparametric Stats by Sex")
```

Days and Duration: Nonparametric Stats by Sex

| Sex | Median Days | 1st Quartile Days | 3rd Quartile Days | Median Duration | 1st Quartile Duration | 3rd Quartile Duration |
|---|---|---|---|---|---|---|
| Female | 135 | 58.50 | 361.00 | 4 | 1 | 11 |

| Sex | Median Days | 1st Quartile Days | 3rd Quartile Days | Median Duration | 1st Quartile Duration | 3rd Quartile Duration |
|---|---|---|---|---|---|---|
| Male | 163 | 59.75 | 431.25 | 7 | 1 | 18 |

2. Make a table of parametric statistics (mean and SD) for the performance outcomes `piq` and `viq` grouped by `sex`. Like the table above, you'll be using `summarise()` from the `dplyr` package. Now you'll use `mean(x, na.rm=TRUE)` and `sd(x, na.rm=TRUE)`. Give the table a title using the `caption=` option and update the column names with something nice using the `col.names=` option in the `knitr::kable()` command.

# Making a table of parametric statistics (mean and SD) for the performance outcomes `piq` and `viq` grouped by `sex'

```
T2 <- Wong %>%
    group_by(sex)%>%
    summarise(meanpiq= mean(piq, na.rm=TRUE),
              stddevpiq= sd(piq, na.rm=TRUE),
              meanvic= mean(viq, na.rm=TRUE),
              stdvic= sd(viq, na.rm=TRUE))
knitr::kable(T2,
             col.names=c("Sex","Mean PIQ","SD PIQ","Mean VIQ","SD VIQ"),
                 caption="Performance Outcomes Piq and Vic: Parametric Stats by Sex")
```

Performance Outcomes Piq and Vic: Parametric Stats by Sex

| Sex | Mean PIQ | SD PIQ | Mean VIQ | SD VIQ |
|---|---|---|---|---|
| Female | 89.18310 | 17.99866 | 94.35211 | 14.24690 |
| Male | 87.11154 | 14.25658 | 95.13077 | 14.02281 |

3. Make a table containing the frequencies and relative percentages for `agegrp`. Use the example we did in class to help guide you.

# Making a table containing the frequencies and relative percentages for `agegrp`

```
# find the sample size or number of rows in dataset
ss <- length(Wong$agefrp)
# group by each level of Medu
# find the raw counts using n()
# then compute the relative % by dividing by ss
T3 <- Wong %>%
  group_by(agegrp) %>%
  summarise(freq = n(),
            pct = n()*100/ss)

# make a summary table
knitr::kable(T3,
             col.names = c("Age Group",
                           "Frequency",
                           "Percent"),
             caption="Frequency Table for Age Group")
```

Frequency Table for Age Group

| Age Group | Frequency | Percent |
|---|---|---|
| Ages 1-10 | 1 | Inf |
| Ages 11-10 | 53 | Inf |
| Ages 21-10 | 144 | Inf |
| Ages 31-10 | 48 | Inf |
| Ages 41-10 | 42 | Inf |
| Ages 51-10 | 27 | Inf |
| Ages 61-70 | 14 | Inf |
| Ages 71-100 | 2 | Inf |

4. Make a regression model (Model 1) for the performance IQ ( `piq` ) using `age` and `sex` . Put the regression model results into a table. ##Making a regression model (Model 1) for the performance IQ ( `piq` ) using `age` and `sex`

```
#run the model for piq with sex and age
Mod1 <- lm(piq ~ age + sex, data= Wong)

# look at the results
SM1 <- summary(Mod1)
knitr::kable(SM1$coefficients,
             caption="Model 1 : piq~age+sex")
```

Model 1 : piq~age+sex

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 86.8868114 | 2.5796740 | 33.681314 | 0.0000000 |
| age | 0.0743977 | 0.0600527 | 1.238874 | 0.2162781 |
| sexMale | -2.1651531 | 2.0258169 | -1.068780 | 0.2859546 |

5. Make a second regression model (Model 2) for performance IQ ( `piq` ) using `age` and `sex` plus `days` and `duration` . Put the regression model results into a table.

```
#run the model for piq with sex and age
Mod2 <- lm(piq ~ age + sex + days + duration, data= Wong)

# look at the results
SM2 <- summary(Mod2)
knitr::kable(SM2$coefficients,
             caption="Model 2 : piq~age+sex+days+duration")
```

Model 2 : piq~age+sex+days+duration

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 88.0961373 | 2.6462149 | 33.2913764 | 0.0000000 |
| age | 0.0542142 | 0.0604989 | 0.8961181 | 0.3708509 |

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| sexMale | -1.7252891 | 2.0152576 | -0.8561135 | 0.3925638 |
| days | 0.0011534 | 0.0007457 | 1.5468461 | 0.1228705 |
| duration | -0.1026657 | 0.0328189 | -3.1282468 | 0.0019172 |

6. Finally, make a table showing the results from the `anova()` command comparing Model 1 and Model 2 you made above using the example we did in class as a guide.

## Compare models using the `anova()` function

```
anova(Mod1,Mod2)
```

```
## Analysis of Variance Table
##
## Model 1: piq ~ age + sex
## Model 2: piq ~ age + sex + days + duration
##   Res.Df   RSS Df Sum of Sq      F  Pr(>F)
## 1    328 74968
## 2    326 72586  2    2381.9 5.3489 0.00518 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Put comparison in a table

```
M1M2 <- anova(Mod1,Mod2)
row.names(M1M2) <- c("Model 1","Model 2")
knitr::kable(M1M2,
             caption = "Comparing Mod1 and Mod2")
```

Comparing Mod1 and Mod2

| | Res.Df | RSS | Df | Sum of Sq | F | Pr(>F) |
|---|---|---|---|---|---|---|
| Model 1 | 328 | 74967.59 | NA | NA | NA | NA |
| Model 2 | 326 | 72585.66 | 2 | 2381.933 | 5.348922 | 0.0051796 |

7. STUDENT CHOICE - pick either a `htmlwidget` from http://gallery.htmlwidgets.org/ (http://gallery.htmlwidgets.org/) or do a “flexdashboard” using the templates at http://rmarkdown.rstudio.com/flexdashboard/ (http://rmarkdown.rstudio.com/flexdashboard/) as a guide.

# Making an interactive datatable with the ANOVA

```
DT::datatable(M1M2, options = list(pageLength = 2, caption="Comparing Mod1 and Mod2"))
```

**Show** 2 **entries** **Search:**

| | Res.Df | RSS | Df | Sum of Sq | F | Pr(>F) |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| Model 1 | 328 | 74967.5898509986 | | | | |
| Model 2 | 326 | 72585.6571162971 | 2 | 2381.93273470152 | 5.34892224140485 | 0.00517957386964677 |

Showing 1 to 2 of 2 entries Previous 1 Next

Repository at: https://github.com/valerie-mac/N741interactive (https://github.com/valerie-mac/N741interactive)
https://github.com/valerie-mac/N741interactive (https://github.com/valerie-mac/N741interactive)

# References

Wong, P. P., Monette, G., and Weiner, N. I. (2001) Mathematical models of cognitive recovery. Brain Injury, 15, 519–530.

Fox, J. (2016) Applied Regression Analysis and Generalized Linear Models, Third Edition. Sage.